Weiwei Zhang

101 Howard St, San Francisco, CA 94115
(310) 729-1915 • wzhang41@dons.usfca.edu

EDUCATION

Master of Science, Analytics (Data Science)
University of San Francisco, San Francisco, CA   07/2015 (expected)
Coursework: Advanced Machine Learning, Times Series Analysis, Distributed Computing

Bachelor of Science, Mathematics/Applied Science
University of California, Los Angeles, Los Angeles, CA   09/2012
Coursework: Probability & Statistics, Linear Algebra, Stochastic Process

SKILLS

Machine Learning: Python (scikit-learn, SciPy, nltk) and R (caret, e1071, rpart)
Data Acquisition & Processing: Python (BeautifulSoup, Pandas, Psycopg2, PyMongo)
Data Visualization: D3.js, Python (matplotlib), R (ggplot2, animation)
Databases: MySQL, PostgreSQL, MongoDB

EXPERIENCE

Data Scientist Intern   12/2014 - present
Zephyr Health Inc, San Francisco, CA
- Created training datasets using tf-idf scores of ngrams as features.
- Applied a Random Forest Classifier as a feature selection method.
- Predicted the disease area relevancy for medical journals using Logistic Regression.

Data Analyst   07/2013 - 05/2014
City of Hope, Los Angeles, CA
- Identified duplicate records in client management system using SQL queries.
- Analyzed donation data for resource allocation in marketing campaigns.

Market Researcher   08/2012 - 05/2013
LGen Partners, Inc, Los Angeles, CA
- Represented clients in financial services to conduct B2B market research analysis.
- Worked closely with account managers to build business partnerships among companies.

PROJECTS

Kaggle Competition: Facial Keypoints Detection (Python, AWS)   02/2015 - 03/2015
- Applied Neural Network models to locate facial keypoints on images.
- Cross-validated the models using a GPU instance on AWS.

TravelWorldSpunky, connecting people with food and events (Python, R, Shiny) 09/2014
- Acquired restaurant and event data via REST API, RSS feed and web scraping.
- Created an app on Shinyapps.io to share visualizations of acquired data.

Machine Learning on the Waveform Database Generator Dataset (Python)   12/2014
- Applied Gaussian Naïve Bayes and Random Forest algorithms to classify waveforms.
- Conducted exploratory data analysis, feature selection, and overfitting detection.

Time Series Analysis on U.S. Inflation Rate from 1914 to 2014 (R)   11/2014 - 12/2014
- Applied a Box-Jenkins model with volatility clustering on U.S. inflation rate data.
- Applied stationarity & white noise tests, classification theorems, ARCH/GARCH effect test.